

Towards Observable Urban Visual SLAM

by

Li Yang Liu

A thesis submitted in partial fulfilment of the
requirements for the degree of Doctor of Philosophy

Supervisor: Associate Professor Shoudong Huang

Co-supervisor: Professor Gamini Dissanayake

at the

Centre for Autonomous Systems

Faculty of Engineering and Information Technology

University of Technology Sydney

March 2020

Certificate of Original Authorship

I certify that the work in this thesis has not previously been submitted for a degree nor has it been submitted as part of requirements for a degree except as fully acknowledged within the text.

I also certify that the thesis has been written by me. Any help that I have received in my research work and the preparation of the thesis itself has been acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

Signed:

Production Note:

Signature removed prior to publication.

Date: 12/02/2020

Towards Observable Urban Visual SLAM

by

Li Yang Liu

Abstract

Visual Simultaneous Localisation and Mapping (V-SLAM) is the subject of robot state and environment map estimation by drawing inference on camera captured data. It has been a major branch of research and popular in application owing to the rich information and low cost in vision measurement acquisition. However, for applications in urban environments, where the camera-mounted vehicle moves along a straight line direction towards the road scene, a large number of features suffer difficulty in depth estimation due to their small parallax angles, as a result the classical V-SLAM algorithm encounters instability and the system state is often unobservable.

This thesis addresses the issue of Urban SLAM observability associated with monocular cameras. It proposes a novel Bundle Adjustment (BA) formulation that addresses the problem from a fundamental approach – by parameterising map points in an on-manifold ray parallax form the SLAM formulation has a stable configuration that guarantees local state observability despite of presence of low parallax features.

V-SLAM is known to be highly non-convex from its projective image formation principle. Slight off-optimal initial values easily lead to sub-optimal final state estimates. In Urban SLAM this is further exacerbated by collinear camera motion that causes ambiguity in initial state estimation. A robust initialisation method is proposed in this thesis to provide unique near-optimal initial estimates effectively addressing collinearity issues.

For practical use of our algorithm, we demonstrate how the urban scene friendly V-SLAM algorithms are integrated into a real-time Visual Inertial Navigation system (VINS).

A series of quantitative analyses are performed on a few benchmark datasets, demonstrating effectiveness of our algorithm in urban environments.

Acknowledgements

Throughout the last 4 years of research work that became the basis of this dissertation, I received a great deal of support and assistance. I would like to thank my supervisors A/Prof. Shoudong Huang, Prof. Gamini Dissanayake and Dr Liang Zhao for their guidance and support over the last four years.

It is worth noting that playing multiple roles of being a PhD student as well as a wife and mother of two young children has been an extremely challenging experience. I would like to thank my mother for looking after my children with total dedication, thank my deceased father for influencing the knowledge loving personality in me, thank my husband for providing support to fulfill my childhood dream, thank my children for putting up with a busy mum, thank my sister for listening to my anguish and thank my school time teachers for guiding me through the many wonders of mathematics. Also a big thank you to my church Corner Stone Concord which provided great sympathy and love during my difficult times.

Thanks to Dr Gavin Paul, Prof Dikai Liu for encouraging me to embark on the PhD course. I acknowledge Teng Zhang, my PhD classmate made contributions to some of the ideas presented in this thesis. Thanks to Ritesh Kumar for developing the VINS sensing platform with me for his capstone. Many thanks to fellow PhD classmates Jonathan Woolfrey and Liye Sun in CAS-UTS for your valuable discussions and open seminars. My sincere gratitude towards Dr Rory Niland, an ex-colleague and well-respected Physicist from whom I first learned SVD and Mark Jeffree the boss from previous work who taught me Signal Processing.

I acknowledge that this research is supported by an Australian Government Research Training Program Scholarship.

Contents

Declaration of Authorship	iii
Abstract	v
Acknowledgements	vii
Table of Contents	ix
List of Figures	xiii
List of Tables	xv
Nomenclature	xxi
1 Introduction	1
1.1 From Projective Transformation to Structure from Motion	2
1.2 Application of Visual SLAM	3
1.2.1 Stereo V-SLAM	4
1.2.2 RGB-D SLAM	4
1.2.3 Visual Inertial Navigation Systems	5
1.3 Motivation	5
1.4 Contributions	7
1.5 Publications	7
1.5.1 Directly Related Publications	7
1.5.2 Related Publications	8
1.6 Awards	8
1.7 Thesis Outline	9
2 Preliminaries	11
2.1 Brief SLAM Overview	12
2.1.1 The Essence of the SLAM Problem	12
2.1.2 The Challenges of the SLAM Problem	15
2.2 Full SLAM: a Non-Linear Optimisation Problem	17
2.2.1 The Gauss Newton Framework	18

2.2.2	Graph Based SLAM	19
2.2.3	Observability and Local Observability	23
2.2.4	Various Forms of Gauss Newton Owing to Local Observability	24
2.3	Structure from Motion and Bundle Adjustment	25
2.4	Urban V-SLAM Issues	27
2.5	The Manifold Representation	28
2.5.1	Gauss Newton on Manifold	29
2.5.2	Some Mathematical Properties of On-Manifold Operations	30
2.6	Summary	33
3	Parallax Bundle Adjustment on Manifold	35
3.1	Urban Scene Issue	36
3.1.1	Feature Parameterisation as a Remedy	37
3.2	PMBA Feature Parameterisation	39
3.3	State Variable Retraction in Manifold	41
3.4	Error Function and Optimisation Formulation	44
3.5	Theoretical Analysis on Convergence Properties	46
3.6	Evaluation on PMBA Performance	51
3.6.1	Simulation	51
3.6.2	Large Dataset Test	52
3.7	Summary	56
4	Robust Global SfM	59
4.1	Background on Global SfM	60
4.1.1	Degeneracy in Position Estimation due to Collinear Motion	61
4.2	Background on Position Initialisation	61
4.2.1	Stability and Robustness in Position Estimation	62
4.2.2	Graph Representation in Position Estimation	63
4.3	PMBA Compatible Initialisation Overview	64
4.4	Epipolar Pair and Maximum Spanning Tree Formation	64
4.4.1	Epipolar Geometry	65
4.4.2	Maximum Spanning Tree	66
4.5	Orientation and Feature Initialisation	67
4.5.1	Orientation Initialisation	67
4.5.2	PMBA Feature Initialisation	68
4.6	Position Initialisation	68
4.6.1	First Position Registration Attempt: Constrained Least Squares + Non-Linear Optimisation	70
4.6.2	Corrected Position Registration: Linear Programming Optimisation	71
4.6.3	Theoretical Analysis	74
4.6.4	Discussion	75
4.7	Experiments	76
4.7.1	Evaluation on Robust Initialisation Performance	76
4.7.1.1	Simulation Test on Global Initialisation	76

4.7.1.2	Global Initialisation Experiments on Benchmark Datasets (Sequential)	77
4.7.2	Evaluation of SfM Pipelines on Large-Scale Datasets	78
4.8	Summary	80
5	Case Study: Real-time Visual-Inertial Navigation	85
5.1	V-SLAM in Real-Time Systems	86
5.2	Marginalisation in a Sliding-Window Optimiser	87
5.2.1	Schur's Complement	89
5.2.2	First Estimate Jacobians	89
5.3	Visual Inertial System Background	90
5.3.1	Literature Review on VINS	91
5.3.2	Inertial Factor – Pre-integration	94
5.3.3	The VINS Problem Statement	94
5.3.4	Marginalisation in VINS	96
5.4	Integration of PMBA in VINS: with Body Centred Coordinate System	97
5.4.1	Motivation	97
5.4.2	Modification of PMBA for VINS	97
5.5	Benchmark Dataset Tests on PMBA-VINS	100
5.6	The Design of VINS Sensing and Mapping Platform	101
5.6.1	Self Collected Datasets and Tests	108
5.7	Summary	111
6	Conclusion and Future Work	113
6.1	Local Observability via Feature Paramterisation	114
6.1.1	Concluding Remarks	115
6.1.2	Future Work	116
6.2	Ambiguity-Free Initialisation for Collinear Motion	116
6.2.1	Concluding Remarks	116
6.2.2	Future Work	117
6.3	VINS Observability	117
6.3.1	Concluding Remarks	117
6.3.2	Future Work	117
	Appendices	118
A	Jacobians in PMBA formation	119
A.1	Jacobian of Observation Ray	120
A.1.1	Special Factor m_j , Current Pose is Main Anchor	121
A.1.2	Special Factor a_j : Current Pose is Associate Anchor	122
B	Jacobians in PMBA-VINS	123
C	IMU Preintegration Factor's Covariance and Error Derivation	125

Bibliography**127**

List of Figures

1.1	Structure from Motion for Unordered Image collection (DAmelio, 2004) . . .	2
1.2	Structure from Motion for Unordered Image collection (Snavely et al., 2006)	3
1.3	Google street view car (Magazine, 2019)	4
1.4	DJI Matrice 600 Drone (DJI, 2017)	6
2.1	SLAM framework: from Maximum Likelihood to Gauss-Newton optimica- tion process	18
2.2	An example actor graph: red circles are pose nodes, green boxes are land- mark nodes. Lines between nodes are factors	20
2.3	Information matrix structure (Engel et al., 2018)	22
2.4	The dogleg update is a combination of Gauss-Newton and gradient descent update steps.	25
2.5	Example results for offline SfM and realtime V-SLAM system	26
2.6	Bundle adjustment - a least squares fitting problem	26
3.1	Feature parameterisation used in conventional BAs	38
3.2	Feature point \mathbf{F}_j anchored by \mathbf{P}_{m_j} and \mathbf{P}_{a_j} with parallax angle θ_j . An arbitrary observing camera is shown at position \mathbf{P}_i . Directions of ray from \mathbf{P}_{m_j} and \mathbf{P}_i are labeled as $\mathbf{R}_{m_j}\mathbf{n}_j$ and $\xi(\mathbf{N}_{j,i})$ respectively, all in global frame.	40
3.3	Retraction of ray \mathbf{n} in main anchor	41
3.4	Camera measurement error	46
3.5	Compare three BA forms in a simulated scene with problem features	51
3.6	Convergence plots for PMBA, PBA, IDP and XYZ	53
3.7	BA comparison test on KITTI-S0 dataset. Existence of collinear features(yellow) cause IDP(brown), XYZ(green) and PBA(magenta) converge to local mini- mum. PMBA(blue) converges to close to ground truth(red), showing perfect immunity to these features.	55
3.8	BA comparison test (cont'd). Dataset: KITTI-S7. Existence of collinear features (yellow dots) cause IDP(brown), XYZ(green) and PBA(magenta) to trap in high cost region and unable to close the loop. PMBA(blue) converges to close to ground truth(red), unaffected by these features.	56
3.9	BA comparison test (cont'd). Dataset: KITTI-S9. Existence of collinear features(yellow) cause IDP(brown), XYZ(green) and PBA(magenta) unable to close the loop. PMBA(blue) converges to close to ground truth(red), unaffected by problematic features.	57

3.10	BA comparison test (cont'd). Dataset: Malaga. Existence of collinear features(yellow) cause IDP(brown), XYZ(green) converge to local minimum. PMBA(blue) and PMBA (magenta) converges to close to ground truth(red), unaffected by problematic features.	57
4.1	Global SfM pipeline: initialisation (purple) + BA (red).	64
4.2	Epipolar constraint	65
4.3	Maximum spanning tree of the fake-pile experiment. Several edges exist between every two nodes, the red edges indicate the essential edges of the maximum spanning tree. Starting from the tree root, initial orientations are estimated by accumulating relative rotations along this path.	66
4.4	Global SfM pipeline: initialisation (purple) + BA (red).	69
4.5	$\mathbf{R}_i \mathbf{v}_{j,i}$ is the measured ray direction, $\lambda_{j,i}$ is the ray length, $\gamma_{j,i}$ is the l_1 residual. PGILP simplifies PMBA into a position estimation problem: $\sum \mathbf{1}^T \gamma_{j,i}$. This error function (Liu et al., 2019) is geometrically meaningful compared to the cross-product based CLS-NonLin error: $(\mathbf{R}_i \mathbf{v}_{j,i}) \times \bar{\mathbf{N}}_{j,i}$ (Liu et al., 2018).	71
4.6	Pose graph showing 10 pose nodes(red), feature nodes (green) and edges. Pose 3-8 are collinear. Red edges are EG-pairs, they are the only source of inputs in Epipolar graph based methods. The feature induced edges (green) are abundant, they constitute constraints in our PGILP initialisation method	77
4.7	Convergence sequence of proposed PGILP camera position initialisation method, on KITTI dataset S5 (full of collinear motion and EG outliers): random to close-to-optimal, PGILP (blue), ground truth (red).	78
4.8	Compare results of different Global Initialisation methods (LUD, WsNonLin, CLS-NonLin and PGILP, all in blue), versus ground truth (red).	79
4.9	Comparing our pipeline reconstruction results (blue) with reference data (red) (cont.)	82
4.9	Comparing our pipeline reconstruction results (blue) with reference data (red) (cont.)	83
5.1	Sensor data flow in VINS	90
5.2	Sliding window filter used in VINS-Mono (Qin et al., 2018)	96
5.3	Compare VINS results on Euroc MH-01 dataset, $Plx_{avg} = 2.5^0$. Ground truth in red, optimised in green, red line segments are loops	101
5.4	VINS tests on MH-02, no-loop, $Plx_{avg} = 1.25^0$	102
5.5	A typical towards scene motion segment in MAV collected benchmark dataset MH-01.	102
5.6	VINS mapping platform: sensor parts	103
5.6	VINS mapping platform: sensor parts - cont'd	105
5.6	VINS mapping platform: sensor parts - cont'd	105
5.6	VINS mapping platform: sensor parts - cont'd	106
5.7	VINS mapping platform: assembly	107
5.8	Comparing pose and map results (blue) with reference data (red)	108
5.9	Mapping platform recorded data: Dr Chau Chak Wing building, UTS	109
5.10	Self collected dataset: University of Sydney, main quadrangle	110

List of Tables

2.1	Different expressions describing uniqueness of problem solution	24
2.2	Three types of Newton-based methods	24
3.1	Various forms of observation ray in this paper	45
3.2	Comparison of $\mathbf{H}_{\mathbf{FF}}$'s condition number during optimisation and final state error for PMBA, PBA and XYZ-BA	51
3.3	Comparison of convergence performance for PMBA, PBA, XYZ-BA, IDP-BA	58
4.1	Pros and cons in Global initialisation schemes	76
4.2	Global initialisation comparison	80
4.3	Complete pipeline comparison: KITTI and BAL	81
5.1	RMSE of VINS odometry poses on Euroc Datasets	100
5.2	Sensor specification	104
5.3	Details of datasets collected from mapping platform	108
A.1	Jacobian for a factor node	120

Acronyms

1D, 2D, 3D	One-Dimensional, two-Dimensional, three-Dimensional
1DSfM	One dimensional outlier removal and structure from motion
BA	Bundle adjustment
CAS	Centre for Autonomous Systems
CLS-NonLin	A two stage position initialisation method: Constrained Least Squares with Non-Linear optimisation
CLT	Central Limit Theorem
CPU	Central Processing Unit
DL	Powell's Dogleg optimisation method
EKF	Extended Kalman Filter
EG	Epipolar Geometry
g^2o	General Graph Optimisation
GD	Gradient descent
GI	Global initialisation
GN	Gauss Newton
GPS	Global Positioning System
GPU	Graphics Processing Unit

GT	Ground Truth
LDL	A variant of Cholesky decomposition $\mathbf{A} = \mathbf{LDL}^T$, \mathbf{L} for lower unit triangular, \mathbf{D} for diagonal
LM	Levenberg-Marquardt algorithm for non-linear least squares regression problems
LS	Least Squares
LUD	Least Unsquared Deviation position initialisation method
IDP	Inverse depth parameterisation of feature points
IMU	Inertial Measurement Unit
IRLS	Iterative Re-weighted Least Squares
iSAM	Incremental Smoothing and Mapping
MAP	Maximum a posteriori
MAV	Micro aerial vehicle
ML	Maximum-likelihood
MST	Maximum spanning tree
MVG	Multiple View Geometry
NEES	Normalised Estimation Error Squared
NLS	Non-linear Least Squares
PD	Positive definite
PGILP	PMBA global initialisation with Linear Programming
PBA	Parallax Angle based Bundle Adjustment
PMBA	Parallax on manifold bundle adjustment
PMBA+VINS	Visual Inertial Navigation System using PMBA for feature paramtrisation and bundle adjustment

PSD	Positive Semi-definite
QR	QR decomposition, decomposition of matrix \mathbf{A} into a product $\mathbf{A} = \mathbf{QR}$ of an orthogonal matrix \mathbf{Q} and an upper triangular matrix \mathbf{R}
RANSAC	Random Sample Consensus
RGB-D	RGB image with corresponding and depth image
RMS	Root mean square
RMSE	Root mean square error
ROS	Robotic Operating System
S^2	2-sphere: 2-dimensional manifold that can be embedded in Euclidean 3D space.
SD	Steepest descent
SfM	Structure from Motion
SE(3)	Special Euclidean group 3D
SIFT	Scale Invariant Feature Transform
SLAM	Simultaneous localisation and mapping
SO(3)	Special orthogonal group 3D
SURF	Speeded-Up Robust Features
SVD	Singular value decomposition
SWF	Sliding Window Filter
UTS	University of Technology, Sydney
VINS	Visual Inertial Navigation System
VINS-Mono	A realtime VINS opensource software package using monocular camera inputs

VIO	Visual Inertial Odometry
VO	Visual Odometry
WsNonLin	Non-Linear position initialisation method, by Wilson
V-SLAM	Visual SLAM
XYZ	3D Euclidean parametrisation of feature points

Nomenclature

General Notations

\mathcal{X}	A generic symbol for robot state variable in the Manifold domain. Can refer to pose \mathcal{T} and feature \mathcal{F} .
\mathcal{T}_i	the i 'th camera pose, $\mathcal{T}_i = (\mathbf{R}_i, \mathbf{P}_i) \in \mathbb{SE}(3)$
\mathcal{F}_j	Feature j 's PMBA parameters in manifold domain
\mathbf{F}_j	Feature j 's coordinates in Euclidean space
\mathbb{T}_j	Indices of all camera poses observing feature j Distributed according to
\approx	approximately equal
$\xi(\cdot)$	the normalisation operator: $\xi(\mathbf{x}) = \frac{\mathbf{x}}{\ \mathbf{x}\ }$ gives vector \mathbf{x} 's direction.
$\pi(\cdot)$	the homogeneous normalisation operator: $\pi\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}^T\right) = \begin{bmatrix} x/z \\ y/z \\ 1 \end{bmatrix}^T$.
\mathcal{V}	Manifold variable displayed in calligraphic font
\mathbf{V}	Euclidean variable displayed in roman font
Subscript $^{(l)}$	indicates the reference frame is local
$[\dots]^T$	Transpose
$ \dots $	Transpose
$\ \cdot\ $	Euclidean norm
\mathbf{I}_n	Identity matrix of size n
$\mathbf{0}_n$	Zero matrix of size n
\mathcal{N}	Normal distribution.
$O(x)$	Big O notation.
\mathbb{R}^2	The set of points in 2D Euclidean space

\mathbb{R}^3	The set of points in 3D Euclidean space
\mathcal{M}	The Manifold domain
\mathbf{z}	Measurement vector
$f(\cdot)$	Measurement function
\mathbf{H}	Hessian matrix and information matrix
\boxplus	Retraction operator
$[\mathbf{w}]_{\times}$	Anti-symmetric matrix based on 3×1 vector \mathbf{w} , see Section 2.5.2 for definition
∂	Partial derivative
\mathbf{J}_B^A	Jacobian of variable A with respect to variable B
\times	cross product
$\tilde{\mathbf{x}}$	Noisy version of \mathbf{x}
$\bar{\mathbf{x}}$	Perfect estimate of \mathbf{x}